Accelerometer-based Human Activity Recognition and the Impact of the Sample Size

ADAM HARASIMOWICZ, TOMASZ DZIUBICH, ADAM BRZESKI
Gdansk University of Technology
Faculty of Electronics, Telecommunications and Informatics
Department of Computer Architecture
11/12 Gabriela Narutowicza Street, 80-233 Gdansk
POLAND
haras.adam@gmail.com, {dziubich, brzeski}@eti.pg.gda.pl

Abstract: The presented study focused on the recognition of eight user activities (e.g. walking, lying, climbing stairs) basing on the measurements from an accelerometer embedded in a mobile device. It is assumed that the device is carried in a specific location of the user's clothing. Three types of classifiers were tested on different sizes of the samples. The influence of the time window (the duration of a single trial) on selected activities and methods was investigated. A comparison with existing methods from the literature is presented.

Key–Words: context awareness, event/activity detection, pervasive computing, user activities modelling, mobile communication, sensors, smart phones

1 Introduction

The development of the systems recognizing human actions basing on electronic sensors is related to the area of ubiquitous processing and machine learning. The problem has been studied since the late nineties, when Foerster et al. published their work[4, 12]. Applications of the considered systems include medicine, sport or defence-related issues[6, 8, 9]. The systems are referred to as Human Activity Recognition (HAR) systems. The idea behind it is to keep low computational complexity, which does not require server-side processing and at the same time provide a high activity recognition effectiveness, without any user-dependent calibration. Additional requirements imposed on the systems are handling both simple and complex user actions, low power consumption, ease of use and finally - effectiveness not only in laboratory but also in real environment. Selection of the input sensor (or a set of sensors) for capturing the signals related to human activity depends on the specific solution. The most commonly used sensors in HAR systems are accelerometers. Their advantages, as shown in[33], include relatively low power consumption comparing to other sensors such as a microphone or a GPS receiver. Other sensors mentioned in the literature are light sensors, barometers, thermometers, compasses and Moisture infra-red sensors, however, the most useful data for activity recognition are provided by accelerometers and microphones[5]. Another criterion for classification of considered systems is the destination of application. Systems can be designed as stand-alone solutions or as extensions to existing systems. In this work we focus on the second category, that is extending devices with HAR functionality. The most popular devices are smart-phones due to their ease of use[7]. Notable examples of such systems are Kwapisz et al.[8], Ravi et al.[9], Back et al.[1], Diatraea[3], The Hearing Trousers Pocket[2], Scott at al.[11], iLearn of the iPhone[10], MSP[5] and Dernbach et al.[6], which are presented in more detail in the following section. The systems vary in the number of recognized activity classes, types of used classifiers, the process of learning and the recognition effectiveness.

In our study we focus on the extended set of activities (7 common user activities considered in aforementioned works supplemented by the turning activity). In addition to examining the classification methods utilised in the works, we also evaluated the influence of the time window size of a sample on the recognition effectiveness. Studies of this kind have not been conducted yet. It should be emphasized that the size of the time window corresponds to the time of a certain, basic movement within an activity. This time depends not only on the activity, but also on the person who is performing the action.

In the first part of the article selected methods from the literature are presented. Next section covers the test environment and a description of the experiment. Finally, the obtained results are presented and discussed.

ISBN: 978-960-474-379-7

2 Background

Kwapisz et al.[8] used mobile devices running Android system. The data was acquired with a three-axis accelerometer. In the data gathering process twenty nine people were engaged for performing measurements under the supervision of the scientists. The set of recognized activities consisted of running, walking, walking upstairs and downstairs, sitting and standing still. During the experiment, the participants were carrying the phone in the front pocket of the trousers. The sampling rate was constant and set to 20 measurements per second. The measurements of a single activity were divided into ten seconds parts, implying that each part contained 200 measurements. The feature vector consisted of the mean value, the standard deviation, the average distance from the mean and the average time between the largest deviations - the values were calculated separately for each axis. In addition, the vector included the average resultant acceleration and the distribution of values for each axis. For the classification task the authors used a decision tree, a logistic regression and a multi-layer perceptron network. Learning and evaluation involved tenfold cross-validation. The highest results were obtained for the perceptron network and the effectiveness reached 91.7%. For all of the methods the proper recognition of walking on the stairs turned out to be the most difficult.

Ravi et al.[9] for collecting the data used a HP iPAQ mobile device supplemented by a prepared research system consisting of an accelerometer, a battery and a Bluetooth transmitter. The purpose was to recognize a similar set of actions as Kwapisz did, but additionally it was extended with brushing teeth and vacuuming. Notably, during the measurements the data from the first and the last ten seconds were discarded. It was done in order eliminate the samples, which could possibly be incorrectly labelled, as the data labelling was performed in a semi-automatic manner - the time slots, in which the given activity was made were determined by using a stopwatch, and then labelled by a script. The data was acquired at the rate of 50 samples per second. In addition, the measurements where grouped into windows of 256 samples with a 50% overlap between adjacent windows, so that the second half of the current part of data was also present in the following part. The computed features included the mean value, the standard deviation and the energy for each axis, and the correlations between them in pairs. The feature vector was then passed to the classifier. The authors utilised tables and decision trees, k-nearest neighbours classifier, SVM and naive Bayesian classifier. The best results were obtained for a voting setup, which achieved the overall prediction correctness of over 90%. The exception was the case when the learning and test sets were collected by different persons, where Boosted SVM achieved highest effectiveness of 73.33%.

Back et al.[1] modelled a more general case - the positions of the sensors were not fixed and several locations were allowed (pockets in trousers or a shirt, clipped to a belt, backpack or even held in hand). In order to reduce the impact of the location of the phone the second order Butterworth filter was used. In this case, however, the number of recognized user actions was limited to three - standing still, walking and running. Sampling was performed at the rate of 100 scans per second and the data was grouped in 2-second time windows. Therefore each data packet consisted of 200 measurements. Both in the supervised and unsupervised measurements the effectiveness of the presented method reached 100% for recognition of standing and walking and 95% for running.

The authors of DiaTrace system[3], apart from activity recognition, put also effort in assuring ease of use. The provided methods and algorithms enabled recognition of such activities as walking, jumping, running, cycling and driving a car. Additionally considered activities were resting and being active in a different way than mentioned above. The first included sitting and sleeping, while the second could be related to activities such as cleaning, gardening or others. The accelerometers used in the experiments had 20Hz sampling rate. However, the actual sampling was dependent on the rate of acceleration changes. Therefore, the read data were later pre-processed by applying interpolation. DiaTrace enabled identification of the activity with the effectiveness exceeding 95% in the case of locating the phone in the front pocket of trousers. For other device locations system capabilities to correctly identify the activities decreased, for example riding a bicycle was confused with driving a car.

The Hearing Trousers Pocket project[2] is a continuation of DiaTrace system[3]. The system increased the effectiveness of the activity recognition and also extended the set of recognized activities. To achieve this, a microphone was used in order to record sounds from surrounding environment. The study also investigated the impact of the mobile device's location on the amplitude of the sound captured by the microphone. The value of 100% was assigned to the signal amplitude received from a source within a distance of 20 cm from the phone without any obstacles. The phone placed in a bag or a backpack received the signal magnitude of 66%. In contrast, the mobile device located in the trousers pocket was able to receive a signal at the level of 100%. This, however, was possible only if the person carrying the phone remained motionless. Otherwise, the scratching between the microphone and the clothing material introduced significant noise to the signal. For this reason, the possibility of using the microphone was limited. In the first instance it was decided to extend the set of the considered activities with working at the computer, which previously had been interpreted as resting. In this case the microphone was able to record the sound of mouse clicks and keystrokes. In the next step the microphone was used to distinguish between the cycling activity and driving a car. During the recognition process, the extended DiaTrace system generated separate feature vector for each signal, which was then classified by a classifier specific for the signal. Thereafter, the results of the first row of classifiers where processed by another classifier, which made the final decision by considering the context. The latter assumed exclusion of certain activities, which can not directly follow each other. The sampling frequency for the acceleration was 32Hz, and 8kHz for the sound. The chosen classifier was the decision tree due to the high efficiency and relatively low computational complexity. The Data from the sensors were grouped into packets of 4096 samples before the feature vectors were evaluated. Furthermore, the sound signal was considered in the recognition process only if the amplitude was high enough, otherwise the decision was made basing solely on the accelerometer. Frequency spectra of both signals were obtained using a fast Fourier transform. Eventually, the running system ended up using 25% of the CPU. The effectiveness of the recognition combining both an accelerometer and a microphone, has not been stated in the work.

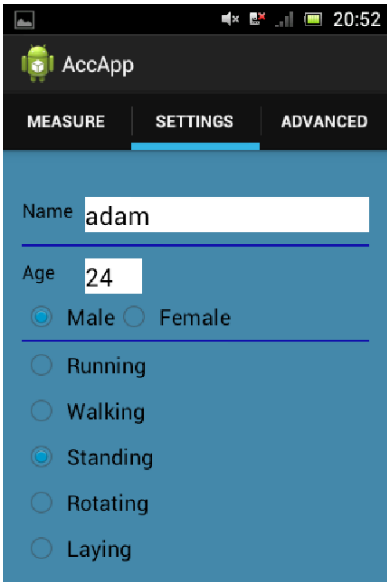

Figure 1: Screenshot of measurement capturing application - AccApp.

Table 1: Number of collected samples for each activity.

Activity	Number of samples
standing	36363
walking	65661
sitting	74584
upstairs	22074
lying	41138
turning	20275
running	34111
downstairs	16991

The iLearn of the iPhone[10] system used an iPhone with a three-axis accelerometer with sampling frequency of 200 Hz and a Nike+ iPod Sport Kit (mounted in the lining of a shoe). The data from the footwear were sent approximately once per second. The Vector consisted of 124 features describing the sample. The obtained features included the mean value, the standard deviation, the minimum and maximum value, the difference between the lowest and highest value and the difference between the lowest and the highest measurement. These values were evaluated separately for each of the three axes within a one-second time window. In order to evaluate the frequency, the discrete Fourier transform was used, calculated every second at 256 points within a time window of 1,25 seconds. The vector also included the energy calculated over the first ten components of the transform and the energy of each component, as well as the highest component's value and index. The values were also determined separately for each axis. For the Nike+ sensor, in turn, full byte stream was put into the feature vector since the device's data format is not documented and the meaning of the data could not be established by observation[10]. In addition, since the Nike+ sensor transmits the data only when the user moves his feet, in case of lack of data the value was set 0. For this reason another feature was also included in the vector, which was set the value of 1 when the Nike+ sensor sent the data in a given time window, and 0 otherwise. Eventually, the average effectiveness of 99.48% was achieved, with the standard deviation of 0.91%. The second experiment assumed selecting the measurements of a single user for the test set, and using measurements of all other user for training the classifier. The average effectiveness obtained in this case reached 97.4% with the standard deviation of 4.05%. The results allowed the authors to conclude, that it is possible to develop a model enabling efficient recognition of activities without a need of training the system in case of appearance of a new user.

ISBN: 978-960-474-379-7

Table 2: Results acquired for each of the classifiers.

Sample size	k-NN		SVM		Decision tree	
	Accurancy	Precision	Accurancy	Precision	Accurancy	Precision
10	98.15%	92.60%	98.45%	93.78%	97.53%	90.12%
20	98.56%	94.23%	98.95%	95.82%	98.09%	92.37%
30	98.70%	94.82%	99.10%	96.42%	98.38%	93.54%
40	98.78%	95.10%	99.23%	96.92%	98.41%	93.65%
50	98.83%	95.32%	99.23%	96.91%	98.42%	93.68%
60	98.87%	95.47%	99.28%	97.12%	98.60%	94.38%
70	98.86%	95.43%	99.19%	96.78%	98.60%	94.39%
80	98.88%	95.51%	99.30%	97.21%	98.48%	93.92%
90	98.78%	95.12%	99.26%	97.03%	98.43%	93.74%
100	98.82%	95.28%	99.27%	97.08%	98.50%	94.00%

3 Test environment and experiments

The analysis of the systems presented above lead us to conducting a study on the impact of the length of the sample considered in the classification on the recognition effectiveness. In the aforementioned works, high effectiveness rates were achieved by various simplifications. Assumption of training the system for each new user, made by some authors, significantly limits its application. Training of the system is necessary if the time of the activities is assumed to be constant (while an older person performs an action slower than a young person), which, unfortunately, is a common case, therefore declining the systems' reliability.

In our work we focused on the data obtained from a smart-phone accelerometer located in the front or side pocket of trousers. The considered activities were walking, running, sitting, lying, standing still, turning, walking up and downstairs. The choice of the set of activities was meant to comply with existing works in order to enable a direct comparison. For the same reason the accelerometer was the sensor chosen. The measurements were collected with the authors' application for data acquisition (AccApp, Figure 1) running on Sony Xperia ST21i with Android Ice Cream Sandwich 4.0.4 and equipped with a three-axis accelerometer Bosch BMA250, with a measurement range of $2g$ (where g is the acceleration due to free fall) and a resolution of approximately $0,0769 m/s^2$. The sampling was performed at the frequency of 10Hz.

The measurements were collected by 8 people on at least 2 different days. In addition, the activities were measured in different environments, which was dependent on the person and the date of the measurement session. Among the participants there were 6 men and 2 women. The youngest participant was 12 years old and the oldest was 53. Each activity was recorded for at least 3 minutes. The total number of collected samples was presented in Table 1.

For further evaluation the following classifiers were selected: k-NN, SVM and decision tree as the most common in the existing systems. After training the classifiers, they were deployed into a mobile application, where the results were recorded for evaluating the system. Two measures were considered: accuracy and precision defined as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$precision = \frac{TP}{TP+FP}$$

where:
TP - True Positive
TN - True Negative
FP - False Positive
FN - False Negative

4 Results and Evaluation

The test procedure included evaluation of recognition effectiveness for the following sample sizes: $\{10, 20, 30, ..., 100\}$. The feature vector consisted of 12 values: the minimum and maximum values, the mean value and the standard deviation calculated for acceleration over each axis. The results where presented in Table 2.

In case of the k-NN classifier the value of k was set to 5, basing on the previous experiments[14]. The highest result was achieved for 80 samples. However, 8 second time window is hard to accept in real environment, since some of the activities (turning, making small steps on the stairs) tend to be performed in much shorter time. Therefore we assume 20 samples as a best solution, resulting in a precision of 94.23%

ISBN: 978-960-474-379-7

Table 3: Classification results for a decision tree with packets of 70 samples. The Ids represent the activities as follows: 1- standing still, 2 - walking, 3 - sitting, 4 - walking upstairs, 5 - lying, 6 - turning, 7 - running, 8 - walking downstairs.

Id	1	2	3	4	5	6	7	8	True	False	Result
1	2555	5	1	0	0	19	0	0	2555	25	99.03%
2	3	4298	43	113	0	3	32	148	4298	342	92.63%
3	9	21	5189	16	26	17	0	12	5189	101	98.09%
4	0	156	26	1165	2	0	0	141	1165	325	78.19%
5	0	0	26	6	2887	0	1	0	2887	33	98.87%
6	22	4	15	0	0	1387	0	2	1387	43	96.99%
7	0	26	8	1	0	0	2381	4	2381	39	98.39%
8	1	178	24	114	0	0	2	801	801	319	71.52%
Total									20663	1227	94,39%

and accuracy of 98.56%.

For the SVM classifier, the best result was also achieved for the packets of 80 samples, but a similar outcome was provided by 60 samples. Nevertheless, again considering the durations of some activities, the number of 20 samples was selected for further experiments. Moreover, none of the 100 iterations (for each of the ten tests, ten-fold cross-validation was performed) resulted in precision rate of 98%.

The tests of the decision tree initially showed an increase of recognition effectiveness along with number of samples. However, starting from the value of 80, the effectiveness dropped. The highest accuracy and precision values were 98.60% and 94.39%, respectively. In Table 3 detailed recognition results of this setup were presented. Analysis of the results leads to a conclusion, that the hardest activities to recognize were walking downstairs, walking upstairs and walking. The three activities were most often confused with each other, which was also observed in the case of k-NN and SVM classifiers. The highest single result was acquired for standing still with 99.03% of properly classified measurement packets.

5 Summary and Future Works

In the presented research we show that the choice of k-NN, SVM and decision tree classifiers allows to achieve high recognition effectiveness. In addition, we show that using only an accelerometer with sampling frequency of 10Hz is sufficient for activity recognition with high accuracy and precision, which is important in terms of limiting power consumption.

The results achieved by the tested methods do not differ significantly. The highest results were achieved for a fairly long time window of seven to eight seconds, which is unacceptable due to the nature of the considered activities. We therefore conclude, that the most reasonable choice is a two second time window, corresponding to a packet of 20 samples. For this value, the precision and accuracy of the SVM classifier reached 95.82% and 98.95%, respectively. We state that using the time window of 2 seconds enables reasonably effective activity recognition with simultaneous non-ignoring of short activities. However, the studies showed that using greater time windows often enables achieving higher results in the test environment. Future work will include extending the system with a solution enabling effective activity recognition regardless of the phone location. Previous studies of this problem resulted in either low effectiveness or a significant reduction of the number of recognized activities.

References:

[1] J. Yun, Recognition of User Activity for User Interface on a Mobile Device, *Special Issue of the International Journal of the Computer, the Internet and Management*, 15(SP4), 2007.

[2] G. Bieber, A. Luthardt, C. Peter and B. Urban. The hearing trousers pocket: activity recognition by alternative sensors, *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA 11*, pages 44:144:6, New York, NY, USA, 2011, ACM.

[3] G. Bieber, J. Voskamp and B. Urban, Activity recognition for everyday life on mobile phones, *Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments*, UAHCI 09, pp. 289296, Berlin, Heidelberg, 2009, Springer-Verlag.

[4] M. F. A. Abdullah, A. F. P. Negara, M. S. Sayeed, D. Choi and K. S. Muthu, Classification Algorithms in Human Activity Recognition using Smartphones, *International Journal of Computer and Information Engineering*, 6, 2012.

[5] T. Choudhury, G. Borriello, S. Consolvo, D. Haehnel, B. Harrison, B. Hemingway, J. Hightower, P. Klasnja, K. Koscher, A. LaMarca, J. A. Landay, L. LeGrand, J. Lester, A. Rahimi, A. Rea and D. Wyatt, The mobile sensing platform: An embedded activity recognition system, *IEEE Pervasive Computing*, 7(2):3241, 2008.

[6] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas and D. J. Cook, Simple and complex activity recognition through smart phones, *Intelligent Environments12*, pp. 214221, 2012.

[7] Top five worldwide smartphone vendors, shipments, and market share calendar year 2011 (units in millions), *IDC*, http://www.idc.com/getdoc.jsp? containerId=prUS23299912, luty 2012. [accessed July 2012]

[8] J. R. Kwapisz, G. M. Weiss and S. A. Moore, Activity recognition using cell phone accelerometers, *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data*, pages 1018, 2010.

[9] N. Ravi, N. Dandekar, P. Mysore and M. L. Littman, Activity recognition from accelerometer data, *Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence (IAAI)*, pages 15411546, AAAI Press, 2005.

[10] T. S. Saponas, J. Lester, J. Froehlich, J. Fogarty and J. Landay, iLearn on the iPhone: Real-time human activity classification on commodity mobile phones. Technical report, 2008.

[11] T. S. Saponas, B. C. Meyers and A. J. B. Brush, Human-activity recognition (HAR) everywhere: Embedded, mobile, desktop, home & cloud, *Pervasive 2011 Workshop: Frontiers in Activity Recognition using Pervasive Sensing (IWFAR)*, Microsoft Research, 2011.

[12] O. D. Lara and M. A. Labrador, A Survey on Human Activity Recognition using Wearable Sensors, http://www.cse.usf.edu/files/69521327941111Survey.pdf, 2012. [accessed January 2013]

[13] Y. Wang, J. Lin, M. Annavaram, Q. A. Jacobson, J. Hong, B. Krishnamachari and N. Sadeh, A framework of energy efficient mobile sensing for automatic user state recognition, *Proceedings of the 7th international conference on Mobile systems, applications, and services, MobiSys 09*, pp. 179192, New York, NY, USA, 2009, ACM.

[14] A. Harasimowicz, Rozpoznanie zachowan uzytkownika na bazie pomiarow z sensorow urzadzenia mobilnego, master thesis, WETI, Gdansk, 2012.